

APPLICATION
FOR
UNITED STATES LETTERS PATENT

APPLICANT NAME: Ravinder Prakash

TITLE: SYSTEM AND METHOD FOR SEGMENTING AN
ELECTRONIC IMAGE

DOCKET NO.: CHA920030015US1

INTERNATIONAL BUSINESS MACHINES CORPORATION

CERTIFICATE OF MAILING UNDER 37 CFR 1.10

I hereby certify that, on the date shown below, this correspondence is being deposited with the United States Postal Service in an envelope addressed to: Mail Stop Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria VA 22313-1450 as "Express Mail Post Office to Addressee" Mailing Label No. EV263593800US

on August 12, 2003

Jennifer E. Desbiens

Name of person mailing paper

Jennifer E. Desbiens 8/12/2003

Signature

Date

SYSTEM AND METHOD FOR SEGMENTING AN ELECTRONIC IMAGE

BACKGROUND OF THE INVENTION

1. Technical Field

[001] The present invention relates generally to a system and method for segmenting an electronic image, and more specifically relates to segmenting electronic images using JPEG DCT coefficient information.

2. Related Art

[002] In order to reduce processing costs, banks and other institutions are seeking to store and process more and more paper documents, including checks, as electronic images. New laws are furthering this movement by, for example, allowing banks to transfer only the check images and associated data for downstream processing. This reduces the requirement for a bank to physically handle and forward paper checks after they are initially received and processed at the bank. However, such scenarios require efficient systems for processing electronic images.

[003] One particular challenge involves capturing information from an electronic image. For instance, in an electronic image of a bank check, the bank may want to identify specific textual or written areas from the check image, such as dollar amounts, signatures, etc. In order to automate this process, foreground elements or segments on the document must be distinguished from the background. Currently there exist numerous solutions in the industry, but most tend to be mathematically complex. For instance, U.S. Patent

5,995,657, "Image Processing Method and Apparatus," issued on November 30, 1999, which is hereby incorporated by reference, utilizes an orthogonal transformation to achieve an image cut out. Such a system requires intensive computations that can substantially increase processing times and costs.

[004] Accordingly, a need exists for a simple and efficient method of segmenting an electronic image to distinguish foreground information from a background region.

SUMMARY OF THE INVENTION

[005] The present invention addresses the above-mentioned problems, as well as others, by providing a system and method for identifying foreground segments in a JPEG compressed image by analyzing DCT frequency data. In a first aspect, the invention provides a method for identifying foreground segments of a JPEG image, comprising: selecting a block in the JPEG image; extracting a set of DCT coefficients from the block, wherein the set comprises the first N AC components of the block; computing a sum of the set of DCT coefficients; and analyzing the sum to determine if the block is part of a foreground segment.

[006] In a second aspect, the invention provides a program product stored on a recordable medium for identifying foreground segments in a compressed electronic image, the program product comprising: means for extracting a set of frequency coefficients from each block of the compressed electronic image, wherein each set comprises the first N frequency coefficients of the block; means for computing a sum of the extracted frequency coefficients for each block; and means for analyzing the sum to determine if the block is part of a foreground segment.

[007] In a third aspect, the invention provides a system for identifying foreground segments of a JPEG image, comprising: a system for extracting a set of DCT coefficients from each block of the JPEG image; a system for computing a sum of the extracted DCT coefficients for each block; and a system for analyzing the sums computed for each block to determine which blocks in the JPEG image are part of a foreground segment.

BRIEF DESCRIPTION OF THE DRAWINGS

[008] These and other features of this invention will be more readily understood from the following detailed description of the various aspects of the invention taken in conjunction with the accompanying drawings in which:

[009] Figure 1 depicts an image segmentation system in accordance with the present invention.

[010] Figure 2 depicts a grayscale image of a bank check.

[011] Figure 3 depicts an image of the bank check of Figure 1 represented using only first DCT coefficients.

[012] Figure 4 depicts an image of the bank check of Figure 1 represented using the sum of AC components of the DCT.

[013] Figure 5 depicts a pair of graphs showing the value differences between an image represented using only first DCT coefficients, and the sum of AC components of the DCT.

DETAILED DESCRIPTION OF THE INVENTION

[014] Referring now to the drawings, Figure 1 depicts an image segmentation system 10 that receives a compressed electronic image 18, and determines one or more identified foreground segments 20. In an exemplary embodiment, compressed electronic image 18 comprises a JPEG image of a grayscale document. However, it should be understood that electronic image 18 may comprise any type of compressed electronic image in which spatial pixel data is represented as frequency information. In the case of JPEG, the frequency information for each block is represented in the form of 63 discrete cosine transform (DCT) coefficient values (DCT[2] ... DCT[64]). The first DCT coefficient value, DCT[1], represents an average video value for the block.

[015] As can be seen, the grayscale document includes a background region 17, and one or more foreground segments 19. In this case, the foreground segments 19 comprise textual data. However, it should be understood that other types of foreground segments, such as graphics, pictures, etc., could likewise be identified. As noted above, it is often desirable to be able to identify foreground segments in a compressed electronic image 18, e.g., for processing purposes.

[016] Image segmentation system 10 includes a DCT analysis system 12 and a foreground identifier system 22. DCT analysis system 12 analyzes the DCT coefficients from each block in a JPEG image and assigns a value to each block, in order to distinguish between foreground segments 19 and the background region 17. DCT analysis system 12 includes DCT coefficient extractor 14, which extracts the first N AC components for each block in the JPEG image, i.e., DCT[2] ... DCT[N+1]. In an exemplary embodiment, DCT coefficient extractor 14 extracts the first sixteen (16)

coefficients from each block. However, it should be understood that the number of extracted AC components can vary without departing from the scope of the invention. DCT coefficient adder 16 adds the magnitude of the extracted DCT coefficients for each block. Thus, a value, referred to herein as a “sum” is calculated for each block of the JPEG image, so that $\text{sum} = \text{DCT}[2] + \text{DCT}[3] \dots + \text{DCT}[N+1]$, wherein all DCT values are first converted to positive numbers.

[017] An exemplary bank check will typically comprise about 5967 blocks (117 x 51 blocks). Accordingly, the bank check can be mapped into 5967 values or sums. Once the sums are calculated for each block, foreground identifier system 22 can examine the sums and distinguish between foreground segments 19 and the background region 17. The identified foreground segments 20 can then be output, stored or otherwise processed as needed.

[018] In general, using the above-mentioned algorithm, a block from a foreground segment 19 will have a significantly greater sum than a block from a background region 17. Accordingly, a relatively simple comparison algorithm may be employed by foreground identifier system 22 to accurately recognize blocks that belong to a foreground segment. For instance, in one embodiment, the sum for each block can be compared to the sums for each neighboring block. If a block is found that has a significantly higher sum than that of a neighboring block, then the block can be tagged as belonging to a foreground segment, and the neighboring block can be tagged as belonging to the background region.

[019] In another embodiment, the sum for each block can simply be compared to a predetermined threshold value. If the sum is greater than the threshold value, then the

associated block can be tagged as a foreground segment. Alternatively, if the sum is less than the threshold value, then the associated block can be tagged as a background region.

As noted above, any methodology can be employed to identify foreground blocks.

[020] Referring now to Figures 2-4, and more particularly Figure 2, a grayscale image of an exemplary check 30 is shown. The check 30 includes foreground segments 34 and a background region 32. To the right of the check 32, is a small snippet 36 that includes the letters “USPS,” cut from the check for exemplary purposes. Figure 3 depicts a representation 38 of the check 30 using an average video value for each block (i.e., $DCT[1]$, the DC value from each block). As can be seen, the foreground segments 40 do not substantially stand out from the background region. This is also evident in the snippet 42, in which it is hard to distinguish between the foreground letters “USPS” and the background region.

[021] Figure 4 depicts a representation 44 of the check 30 using the Image Segmentation System 10 described above. In particular, the representation 44 uses a value for each block obtained by summing the first 16 frequency coefficients for each block (i.e., $DCT[2] \dots DCT[17]$). As is evident, the foreground segments 46 significantly stand out from the background region since the signal to noise ratio for the foreground information in the representation 44 is increased in the frequency domain as compared to the average video domain of Figure 3. This is likewise the case for the snippet 48, in which the foreground information “USPS” can be readily distinguished from the background.

[022] Figure 5 depicts the corresponding information for the snippets 42 and 48 in a pair of three-dimensional bar charts 42’ and 48’. Each bar chart includes 84 values that

represent 14 x 6 blocks required for each snippet. Bar chart 42' corresponds to snippet 42 (average video domain), and bar chart 48' corresponds to snippet 48 (frequency domain). As can be seen, the ability to discriminate between foreground and background blocks is greatly enhanced in bar chart 48', which results from assigning values to blocks utilizing the calculated sums described above. For instance, the difference between a typical background value 50 and foreground value 52 in bar chart 42' is substantially less than the difference between a typical background value 54 and foreground value 56 in bar chart 48'. Accordingly, greater confidence in identifying foreground segments can be achieved by assigning values to blocks using frequency domain components.

[023] It is understood that the systems, functions, mechanisms, methods, and modules described herein can be implemented in hardware, software, or a combination of hardware and software. They may be implemented by any type of computer system or other apparatus adapted for carrying out the methods described herein. A typical combination of hardware and software could be a general-purpose computer system with a computer program that, when loaded and executed, controls the computer system such that it carries out the methods described herein. Alternatively, a specific use computer, containing specialized hardware for carrying out one or more of the functional tasks of the invention could be utilized. The present invention can also be embedded in a computer program product, which comprises all the features enabling the implementation of the methods and functions described herein, and which - when loaded in a computer system - is able to carry out these methods and functions. Computer program, software program, program, program product, or software, in the present context mean any expression, in any language, code or notation, of a set of instructions intended to cause a

system having an information processing capability to perform a particular function either directly or after either or both of the following: (a) conversion to another language, code or notation; and/or (b) reproduction in a different material form.

[024] The foregoing description of the preferred embodiments of the invention has been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise form disclosed, and obviously many modifications and variations are possible in light of the above teachings. Such modifications and variations that are apparent to a person skilled in the art are intended to be included within the scope of this invention as defined by the accompanying claims.